First statement of D. Carroll

On behalf of the Claimant

Exhibits C001 – C232

16 March 2018

IN THE HIGH COURT OF JUSTICE **Claim no. _____**

QUEEN'S BENCH DIVISION

B E T W E E N

DAVID CARROLL

Claimant / Applicant

V

(1) CAMBRIDGE ANALYTICA LTD
(2) CAMBRIDGE ANALYTICA (UK) LTD
(3) SCL ELECTIONS LTD
(4) SCL GROUP LTD

Defendants / Respondents

STATEMENT OF CLAIMANT

I, DAVID CARROLL, of 29 Tiffany Place, Apartment 1K, Brooklyn, New York, NY 11231 STATE as follows in in support of my application for pre-action disclosure:

A. Introduction

1. Save as set out herein, the matters set out in this witness statement are within my knowledge. Where the facts are not within my own knowledge, they are true to the best

of my knowledge, information and belief and I have indicated the source of my information.

B. My background

2. I am an Associate Professor of media design at the Parsons School of Design, New York. I have two decades of professional experience in the digital advertising and product development industries. I was promoted and granted presumed employment status in 2013 by the Parsons Tenure committee in recognition of my academic profile, excellence in teaching, and service to the institution and related fields.

3. My professional and academic experience has allowed me to develop a unique expertise and insight into online behavioural advertising technologies and consumer attitudes about these practices.

4. Through my academic research, coupled with my own entrepreneurial experience of developing data-based media platforms, I am familiar with how the digital media and marketing industry is able to track user behaviours. In particular, I have experience with how personally identifiable information is used by such companies, as well as how they employ "re-identification" techniques and technologies, whereby anonymous data and information are combined with other information to "re-identify" an individual from anonymous data. It is also often described as "de-anonymization" but the industry prefers terms such as "onboarding" and "hygiene (data)" (a phrase the Defendants use[1]) and other jargon that obfuscates the practices of data matching across data sets for blending. The purpose of this practice is to make communications addressable to individuals based on the traces of their personal and aggregate behaviours.

5. Through my academic research, I understand that these methods are troubling to the wider public as these technological developments do not comport with people's expectations of privacy. In particular, these "de-anonymization" techniques do not always comply with the expectations of privacy protection created by typical privacy policies. Such policies generally attempt to assure users that their behavioural data can be safely shared with third-parties once our personal identifiers are removed.

[1] See, bundle page [C189] at q792

6. As a result of my work in this area, I have been invited by the press and digital publishing trade groups to provide commentary and insights regarding the privacy and consumer trust issues surrounding personalization and targeting. [2]

7. It is important for academic researchers, journalists, and policy makers to be able to investigate whether or not re-identification / data-matching is occurring within the industry to better understand the data supply chain. This is especially true for voter analytics, where a voter's consumer behavior may be re-attached to their voter registration.

C. My decision to request my information from Cambridge Analytica / SCL

8. The Defendants are well known and their activities are controversial. They have attracted controversy both for their profiling techniques and for being alleged to have been used by the Trump Presidential campaign. There are numerous reports about the company, such as:

[2] See, for example, *The New York Times*, September 18, 2015, "Enabling of Ad Blocking in iOS 9 Prompts Backlash", Katie Benner and Sydney Ember, https://nyti.ms/1OER8JW; *Mashable*, November 6, 2015, "Ad-blocking is so widely loathed that one company held a peace summit about it", Patrick Kulp, https://on.mash.to/1lzsyzT; *Independent* (UK), June 3, 2016, "Facebook Adds New Options as People Complain That They've Been Signed Up By Default", Andrew Griffin, http://www.independent.co.uk/life-style/gadgets-and-tech/news/facebook-adds-new-ad-options-as-people-complain-that-they-have-been-signed-up-by-default-a7064481.html; *USA TODAY*, June 5, 2016, "Status with Facebook 'interest-based' ads complicated', Rob Pegoraro, http://usat.ly/1sYVqog; *Washington Post*, November 18, 2016, "This Is How Facebook fake news writers make money", Abby Olheiser, http://wapo.st/2g5wJC0; *Digiday*, December 13, 2017, "Why Top Publishers Are Still Stuck Distributing Fake News", Lucia Moses, http://digiday.com/media/top-publishers-still-stuck-distributing-fake-news/; *Associated Press*, December 29, 2016, "Study: Ad-tech use shines light on fringe, fake news sites", Barbara Ortutay, https://apnews.com/5e0e7b1c9ae94c619c31b6fb024ec379; *Business Insider*, February 16, 2017, "30 Best People to in Advertising to Follow on Twitter", Julien Rath, http://www.businessinsider.com/the-30-best-people-in-advertising-to-follow-on-twitter-2017-1/#9-david-carroll-22 ; *The Wall Street Journal* (Dow Jones Newswires), October 29, 2017, ""Who Will Rein in Facebook? Challengers Are Lining Up", Christopher Mims. http://www.foxbusiness.com/features/2017/10/29/who-will-rein-in-facebook-challengers-are-lining-up.html ; *Quartz*, October 22, 2017, "Twitter has a serious bot problem, and Wikipedia might have the solution", Robert Gorwa, https://qz.com/1108092/twitter-has-a-serious-bot-problem-and-wikipedia-might-have-the-solution/ ; *ThinkProgress*, September 19, 2017, "The President of the United States is regularly manipulated by Twitter bots", Luke Barnes and Phoebe Gavin, https://thinkprogress.org/the-president-of-the-united-states-is-regularly-manipulated-by-twitter-bots-2bb5ee8321a8/ ; *The Observer*, September 30, 2017, "British courts may unlock secrets of how Trump campaign profiled US voters", Carole Cadwalladr, https://www.theguardian.com/technology/2017/oct/01/cambridge-analytica-big-data-facebook-trump-voters ; *Slate*, October 6, 2017, "Potential Lawsuit Could Reveal How Trump Targeted Voters on Facebook and If There's Any Connection to Russia" April Glaser, http://www.slate.com/blogs/future_tense/2017/10/06/possible_british_lawsuit_could_reveal_how_cambridge_analytica_targeted_voters.html

i. Hannes Grassegger and Mikael Krogerus for *Motherboard*, "The Data That Turned the World Upside Down" ([exhibit bundle page C008 – C024])
ii. Carole Cadwalladr in the *Observer*, "The great British Brexit robbery: how our democracy was hijacked" ([exhibit bundle page C046 – C059])
iii. Nicholas Confessore and Danny Hakim for *The New York Times*, "Data Firm Says 'Secret Sauce' Aided Trump; Many Scoff" ([exhibit bundle page C028 – C038])

9. My inquiry into Cambridge Analytica processing my personal data originated from academic curiosity and a general desire to ensure that my personal data was not used for purposes I consider to be unsettling or unlawful. This was particularly true with respect to my political opinions.

10. I have a professional understanding of how various sources of my consumer data can be re-attached to my political identity, while I also understand how ad tech, set-top boxes, and social media platforms allow advertisers to "micro-target" audiences. As such, I was concerned that members of my family and myself were being exposed to political messages that attempted to influence our sentiments about particular candidates during the primaries and general elections without a clear understanding of their origination or intent.

11. My academic interests in consumer attitudes about data privacy and the advertising industry's use of personal data for targeting and personalization led me to became more thoroughly acquainted with academic experts in data science and data protection based in Europe, especially Paul-Olivier Dehaye, a Swiss-based mathematician. I had discussed Cambridge Analytica in particular with Dr Dehaye, as I had heard about them through public reports and filings which indicated that it provided significant support to multiple presidential campaigns during the 2016 election cycle, namely the Carson, Cruz, and Trump campaigns. I understood that my data was being processed through Europe and, more particularly, in London. I wished to understand what data was being processed and who it was going to. Dr Dehaye advised me on how to submit a Subject Access Request ('SAR') to Cambridge Analytica to request my personal data from them.

D. The SAR Process

12. I submitted my SAR in or around 10 January 2017. I received a substantive response on 27 March 2017. The results of the SAR confirmed that my US voter files were processed in the UK. When I submitted my SAR, I did not anticipate uncovering facts that have become relevant to various investigations underway in the United States relating to the 2016 Presidential Election. I was distressed to receive the information contained in the SAR from the Defendant companies; information that was deeply personal – even though the disclosure made remains incomplete.

13. The process of submitting a SAR to Cambridge Analytica/SCL was as follows:

    a. I visited the website https://datarequests.cambridgeanalytica.org on 10 January 2017 to initiate the request;

    b. I received a response email from data.compliance@sclgroup.cc with instructions to fulfil the request. I was asked to wire a £10 processing fee to SCL Elections, Ltd. I was also asked to provide photographic identification documents and a utility bill for identification verification purposes. I should add that I was reluctant to provide this personal information, given my concerns about the Defendants. However, the Defendants made providing this information obligatory and so I provided a copy of my driving licence and a utility bill.

    c. The final substantive response was sent on 27 March 2017, under the cover of a letter signed by Julian Wheatland, COO of SCL Group "For and on behalf of Cambridge Analytica"

14. The Profile, as at 27 March 2017, included information in the following categories:

    - Core data – Background information, including my name, address, date of birth and voter ID.
    - Election returns – All data relating to my election returns for both primary and general elections from 2000 – 2014, including who I cast my vote for. The

information concerning my voter registration is not readily available in the public domain. The rules vary from state to state concerning this information but in my state, New York, such information is only available to registered electoral candidates, through an application to the New York State Department of State. Furthermore, the information concerning who I voted for is necessarily unknown and confidential by its nature. However, the information provided by the Defendants (particularly the "eresult" column) is not clear as to whether this is a record of who I voted for or who won the respective vote in each entry.

- Models – The "models" category was a political profile of me. In summary, my political views were categorised in ten variables, ranked by order of perceived importance. In addition, my partisanship was categorised by both my registered and unregistered political preference. Finally, the profile contained a perceived propensity to vote in the 2016 election. In essence, this "model" is a bespoke political profile of me.

15. The Defendant companies had profiled my most intimate political views and beliefs. There was no explanation of how this model was created nor what it was based on. Based on my initial review, it seemed likely that other information about me must be held by Cambridge Analytica as I do not see how they could have generated the "models" solely on the basis of the core and election returns dated provided.

16. For example, the profile / "models" put "gun rights importance" at rank 3. In my view, and that of Dr Stillwell (see further below), this is a counter-initiative conclusion to reach about me based on the basic background data Cambridge Analytica provided. However, it is an apt categorisation, as this is an issue that I give importance to. There would be no way to know that about me from publicly available information, as I am not a member of the NRA, and do not subscribe to any gun magazines etc. As such, I am astonished that this ranking could have been made with some precision, without some further underlying and personal information about me. That the Defendants have this information about me and will not disclose it is disconcerting and distressing.

17. I also do not know what companies or other entities have received this information. I was very troubled to receive this profile and could not understand how such an intricate picture of my life was formed.

18. I accordingly instructed solicitors to correspond with the Defendant companies on my behalf regarding that profile. The statement from my solicitor, Ravi Naik, accompanying this statement sets out the chronology of correspondence with the Defendant companies since that date (see bundle page [B014]). The relevant correspondence is exhibited to that statement. I will not rehearse that chronology herein.

E. Outstanding questions and disclosure required

19. My concerns about the extent of the data held on me, why the profile was created and who it was provided to have limited my ability to understand and identify the issues that would be raised in my subsequent claim.

20. Owing to both my professional experiences and from the publically available information about the company, I believe that information has been withheld from me. The publically available information is voluminous but I provide some specific examples, as follows:

    a. Cambridge Analytica's website (recently amended) previously stated that it uses *"more than 100 data variables to model target audience groups and predict the behaviour of like-minded people"*.[3]

    b. According to Alexander Nix, Cambridge Analytica's chief executive, Cambridge Analytica *"have a massive database of 4-5,000 data points on every adult in America"*.[4] The current Cambridge Analytica website states they have "up to 5,000 data points on over 230 million American voters, we build your custom target audience, then use this crucial information to engage, persuade, and motivate them to act."[5] (Exhibit bundle page C200 – C202])

    c. Executives and employees from Cambridge Analytica have disclosed publicly at least some information about their data sources to the public and journalists. This is confirmed in multiple news articles.[6]

[3] See exhibit bundle page C001 – C004.
[4] https://www.ft.com/content/bee3298c-e304-11e6-9645-c9357a75844a
[5] https://ca-political.com/ca-advantage
[6] See, inter alia, https://nyti.ms/2lvpcCj; https://www.bloomberg.com/news/articles/2016-10-27/inside-the-trump-bunker-with-12-days-to-go; http://www.commonwealthy.com/voter-analytics-transcript/;

d. There is a public video of the CEO of Cambridge Analytica, Alexander Nix, from the "Concordia Summit". In that video, he articulates his firm's use of political data. Mr Nix states that the company uses commercial data supplied by third-party vendors to match with political records. That information is then amalgamated to create the profiles.[7] I have exhibited a transcript of the talk by Mr Nix ([exhibit bundle page C005 – C007]). Furthermore, in this video, Mr Nix specifically illustrates how psychographics were employed to measure and target US voters by their personality types and susceptibilities to personalized messages.

e. On 5 September 2017, the BBC released an episode from the series, *Secrets of Silicon Valley* with Jamie Bartlett.[8] In this particular episode, Alexander Nix acknowledges that *"legacy data"* and *"psychographic models"* used during the Ted Cruz campaign were *"carried over into the Trump campaign."*

f. On 27 February 2018, Alexander Nix gave oral evidence before the Digital, Culture, Media and Sport Committee. I exhibit a copy of the transcript of the hearing before the Committee (exhibit page [C153 – C199]). In that Committee hearing, Mr Nix stated: *"In a country such as the United States, we are able to commercially acquire large datasets on citizens across the United States—on adults across the United States—that comprise of consumer and lifestyle data points. This could include anything from their hobbies to what cars they drive to what magazines they read, what media they consume, what transactions they make in shops and so forth. These data are provided by data aggregators as well as by the big brands themselves, such as supermarkets and other retailers."*[9] Further, Mr Nix confirmed that the Defendants use data to create profiles, stating: *"I think I have made my position clear, which is that we are trying to make sure that we can use data to understand what people care about and we can seek to address those concerns"*[10] He went on to state that: *"... All we have done is gone*

https://www.forbes.com/sites/stevenbertoni/2017/05/26/jared-kushner-in-his-own-words-on-the-trump-data-operation-the-fbi-is-reportedly-probing/; https://www.nybooks.com/articles/2017/04/20/kahneman-tversky-invisible-mind-manipulators/

[7] https://www.youtube.com/watch?v=n8Dd5aVXLCc See transcription at exhibit bundle page C005 – C007.

[8] http://www.bbc.co.uk/programmes/b091zhtk

[9] At Q679 – Q680

[10] At Q784

to all the vendors, large, medium and small, and taken these data and put them into one database and record."[11] This would suggest that the Defendants do in fact possess significant underlying material on those they profile.

g. Furthermore, during the hearing referred to directly above, Mr Nix confirmed that the Defendants shared data amongst themselves, confirming that they *"transfer data from Cambridge Analytica to SCL."*[12] This confirms my concerns that my personal – and sensitive personal – data is being transferred by the Defendants. It is not clear to whom or where else my data has been passed to.

h. Mr Nix also confirmed that the profiles were used to *"micro target"* individuals. Mr Nix confirmed that the Defendants in the 2016 Presidential election *"ran 4,000 different advertising campaigns—about 1.4 billion impressions."*[13] Mr Nix also confirmed that the Defendants *"rolled out 350,000 to 400,000 surveys a month for the Trump campaign in the United States over a five-month period. These were being done on behalf of the Trump campaign and that was the label of the survey. The fact that we were helping them to gather these data was less relevant."*[14]

21. Mr Nix's testimony before the DCMS Committee has simply served to confirm my concern that the Defendants have harvested copious data sets concerning me and that the data was transferred to other third parties without my knowledge or consent.

22. My concerns about Cambridge Analytica holding intimate data on me have also been increased by the press and publicity surrounding the Defendants' methodologies (such as Mr Nix's testimony before the Concordia Summit), coupled with the post-election controversies surrounding the company's activities (such as seeking data from Julian Assange / Wikileaks through the designated agent of his speaker's bureau, just as it was starting its work for the Trump campaign[15]). In this context, it is all the more important that I should be provided with disclosure to understand what the company was doing. I

[11] At Q792
[12] At Q688 – Q689
[13] At Q832
[14] At Q772
[15] https://www.wsj.com/articles/data-firms-wikileaks-outreach-came-as-it-joined-trump-campaign-1510339346

cannot understand why I have not been provided with the list of the sources of my data, my psychographic scores, and a comprehensive list identifying the recipients of my data.

23. In an effort to better understand what data was missing, I have instructed two experts to prepare reports into the company and political data. Those experts are Dr David Stillwell of Cambridge University, who has conducted a report into the psychographic profiling employed by the company and Professor Philip Howard from Oxford University, who has prepared a report on the use of data in politics. Both of those reports serve to show that significant data is missing.

24. Those reports have confirmed my suspicions that there must be further data to have given rise to the profile. That data has not been provided to me. In an effort to explain that sorts of data that would be used to give rise to these profiles, I have enclosed a report from the data industry researcher, Wofie Christl. That report, *Corporate Surveillance in Everyday Life*, provides a detailed explanation of how disparate and disconnected data is amalgamated to profile individuals in intricate detail.[16] That report does not relate to the Defendant companies but illustrates the vast data needed to be able to profile individuals in the way I have been profiled by the Defendants. In summary, the report confirms that there must be vastly more data held on me to give rise to the profile than that which has been disclosed by the Defendants.

F. <u>What information I am seeking</u>

25. As I outline further below, I am concerned by the transnational voter data processing that the Defendant Companies undertake and its wider implications. The deficient disclosure provided by the Defendant Companies prevents me from understanding the extent to which my rights have been interfered with. Further, the lack of information relating to the wider concerns that have been publically expressed in relation to the activities of the Defendant Companies are exemplified in my case.

26. My specific concerns regarding this missing information are as follows:

[16] http://crackedlabs.org/en/corporate-surveillance

a. First, without a list of the sources of my data, it is not possible for me to understand how my commercial and/or political behavior could have been modelled into a precise profile that could accurately predict my participation and vote in the 2016 presidential election cycle and beyond. I need transparency on the sources of the data in order to understand how my online or other behavior has led to generation of the profile of me. This in turn will allow me to determine the various kinds of political messaging that I might have been and continue to be exposed to on television and online platforms (see further below).
b. Second, I need further clarification as to the source of my voter registration file. In my state, New York, such voter registration information is not readily available (in contrast to say, Ohio, where such information can be readily downloaded). Rather, such information is only available to potential electoral candidates through an application to New York State Department of State. I also do not understand what the "eresult" column refers to and require disclosure of what this means.
c. Third, the deficient disclosure made, coupled with the failure to explain how the processing has been undertaken, means that I do not know how the Defendant companies processed my data in order to create a profile on me. While the technique of using personality modeling, known as psychometrics or psychographics, has earned the most significant press coverage for the company, I am also interested in determining if the Defendants are engaged in the practice of re-identification by understanding whether this process was applied in my case. I also do not know if voter files are used for Cambridge Analytica's commercial clients.
d. Fourth, I also need more specific details as to the re-identification and modelling methods. Alternatively, if Cambridge Analytica / SCL has made data-matching connections between my voter registration file and commercial databases that contain my personally identifiable information enabling a match with 100% confidence, I need to know where such matching has happened and by reference to specified data sources.
e. Fifth, I do not know who has received my data, in particular, who has bought or otherwise had access to the ideological profile generated in respect of me.

27. Without the above information, it is impossible for me to understand how the Defendant Companies arrived at the profile of me that they generated and disclosed in response to my SAR. I need the disclosure which I requested through my SAR, and outlined in the application for pre-action disclosure, in order to understand how and why my data has been processed, and whether I should proceed with a claim for damages and other relief as a result of the actions of Cambridge Analytica / SCL.

28. The failure to comply properly with my SAR / the failure to make pre-action disclosure inhibits my ability to understand what Cambridge Analytica / SCL have been doing with my data, and my ability to conduct appropriate research into the data analytics industry.

29. Furthermore, owing to my understanding of the mechanics of personal data collection, re-identification, and targeting online, it is distressing for me to consider how my own exposure to targeted and personalized media may have influenced my own voting decisions in the primary and potentially depressed my enthusiasm to actively campaign for the general election candidate.[17]

30. In particular, I am concerned that I may have been targeted with messages that criticised Secretary Hillary Clinton with falsified or exaggerated information that negatively affected my sentiment about her candidacy and consequently discouraging me from engaging with the Clinton campaign as a formal or informal volunteer.

31. Having read press reports about the Trump campaign's deliberate attempt to achieve these effects using data provided by Cambridge Analytica, it causes me distress not being able to understand how the results of my SAR indicate how I may have been targeted with media and advertising on television and social media platforms. In this context, I seek redress from the authorities and courts in England as my data was processed by a UK company.

32. In summary, I still do not know the answers to the following key questions:

    a. What is the underlying data that gave rise to the profile?

[17] Green, Joshua and Sasha Issenberg. "Inside the Trump Bunker, With Days to Go" *Bloomberg Businessweek*, October 27, 2016. https://www.bloomberg.com/news/articles/2016-10-27/inside-the-trump-bunker-with-12-days-to-go

b. Where did Cambridge Analytica / SCL get my personal data?

c. How did Cambridge Analytica / SCL process my voter registration file with commercial data vendors? Did the company use the personal information in my voter registration to re-identify me in the databases of other vendors? What data matching methods did it employ and how accurate are the linkages?

d. With whom did Cambridge Analytica / SCL share my profile?

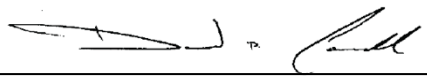
e. With whom were the Defendant Companies liaising with during the election?

f. Did the Defendants obtain and otherwise process my data?

g. How can the Defendants justify creating an ideological model of me without my knowledge or consent?

STATEMENT OF TRUTH

I believe that the facts stated in this witness statement are true.

Signed: 

Named: David Carroll

Dated: 16 March 2018

First statement of R. Naik

On behalf of the Claimant

Exhibits C203 – C234

16 March 2018

IN THE HIGH COURT OF JUSTICE **Claim no._____**

QUEEN'S BENCH DIVISION

B E T W E E N

DAVID CARROLL

Claimant / Applicant

- and -

(1) CAMBRIDGE ANALYTICA LTD
(2) CAMBRIDGE ANALYTICA (UK) LTD
(3) SCL ELECTIONS LTD
(4) SCL GROUP LTD

Defendants / Respondents

STATEMENT OF RAVI NAIK

I, RAVI NAIK, of Irvine Thanvi Natas Solicitors, 5 Stratford Office Village, London, E15 4EA STATE as follows in in support of the Claimant's application for pre-action disclosure:

1. I am a partner and solicitor at Irvine Thanvi Natas Solicitors and I have conduct of the above matter on behalf of the Claimant.

2. Save as set out herein, the matters set out in this witness statement are within my knowledge. Where the facts are not within my own knowledge, they are true to the best

of my knowledge, information and belief and I have indicated the source of my information.

I. BACKGROUND

3. The background of the Defendants are laid out in the witness statement of the Claimant. In summary, the Defendant companies are engaged in behavioural micro-targeting. They are known for the use of such micro-targeting in political campaigns.

4. Owing to concerns surrounding these campaigns, a number of individuals sought the information held about them by the Defendants. Professor Carroll submitted a subject access request to Cambridge Analytica on 10 January 2017, which was processed by the Defendants collectively (as to which, see below). A subject access response was received on 27 March 2017.

5. The Defendants failed to properly comply with that request and provided only limited disclosure. However, the limited disclosure made by the Defendants, taken together with publicly available information about the Defendants, has raised serious concerns from the Claimant about their processing of his personal, including sensitive personal, data. The Claimant has provided a statement on these concerns ([bundle page B001]).

6. We have engaged in pre-action correspondence with the Defendants. Regrettably, we do not have any pre-action document from the Defendants that we can provide to the Court. We address the pre-action correspondence below.

i. The Defendants

7. The Defendants are all based in the United Kingdom. They are all listed within Companies House as operating in the United Kingdom. I exhibit the Companies House listings for the Defendants ([exhibit bundle pages C224 - C226]).

8. There are two companies registered on Companies House with the name "Cambridge Analytica" (Cambridge Analytica Ltd (Co number: 09154503) and Cambridge Analytica (UK) Ltd: 09375920). It is not clear whether they are related companies and, if so, how

these companies operate together. We have sought clarification from the Defendants but we have not been provided a permissible response to show to the Court.

9. SCL Elections Limited[1] is Cambridge Analytica (UK) Limited's parent company[2]. SCL Elections Limited is owned and operated by SCL Analytics Limited[3]. The ultimate parent company of all these subsidiaries is SCL Group Limited, which is registered in the United Kingdom.[4]

10. I have looked at the various Defendants' websites to understand their business. The current SCL Group Limited's website states that it provides "data, analytics and strategy to governments and military organisations worldwide." They further state that they have "conducted behavioural change programs in over 60 countries".[5] ([exhibit bundle page C227 – C228])

11. The websites for the various Defendants have changed since the time of the Claimant's subject access request. In particular, SCL Elections were previously a main page within the SCL Group website but that arm of the company is no longer listed on the website (despite SCL Elections remaining registered at Companies House). I exhibit an earlier version of the website from an archived source which shows that Elections were one of SCL Group's main sectors [6]([exhibit bundle page C207]).

12. At the time of the Claimant's subject access request, the stated commercial aim of the Defendants on the SCL website was to collect data on individuals to facilitate "microtargeting". The previous version of the Cambridge Analytica website further explained that "when you go beneath the surface and learn what people really care about you can create fully integrated engagement strategies that connect with every person at the individual level." ([exhibit bundle page C001 – C004)].

[1] Company number 08256225. SCL Elections Limited are registered with Companies House as the Company with "significant control" over Cambridge Analytica (UK) Limited.
[2] SCL Elections Limited are registered with Companies House as the Company with "significant control" over Cambridge Analytica (UK) Limited.
[3] Company number 09838667
[4] Company number 05514098
[5] http://web.archive.org/web/20180227231636/https://sclgroup.cc/home
[6] http://archive.is/JlkSn

13. According to the previous version of the SCL website, SCL stated that they use "advanced scientific and social analysis techniques to better understand behaviour."[7] ([exhibit bundle page C208 – C209] It is also marketed as being able to "understand the deep attitudes, motivations and social structures of communities in order to influence their long term behaviour and encourage lasting change".[8] ".[9] ([exhibit bundle page C210 – C211]

14. The earlier version of Cambridge Analytica's website, which I had considered when drafting the Claimant's pre-action letter, stated that its database "contains information on demographics, consumer and lifestyle habits and political affiliation, as well as unique psychographic information on motivation and decision-making". I have looked for for the earlier version of the website on internet archives but have been unable to find it. We continue to look for the earlier versions of Cambridge Analytica's website from archived sources.

15. The earlier version of the website also stated that they used "more than 100 data variables to model target audience groups and predict the behaviour of like-minded people" ([exhibit bundle page C001 – C004]). The current Cambridge Analytica website goes on to state that they "source data from a trusted list of major data providers and combine it with your own to produce deeper and richer insights. Then we centralize your data assets and match your data to ours, to help you to find and persuade voters quickly and efficiently."[10] ([exhibit bundle page C229 – C232]). They further state that they have "up to 5,000 data points on over 230 million American voters" and that they use the data to "build your custom target audience, then use this crucial information to engage, persuade, and motivate them to act."[11] ([exhibit bundle page C200 – C202])

16. There has been no public explanation as to the basis for the reasons for the change in the SCL / Cambridge Analytica website.

[7] http://web.archive.org/web/20170219092643/https://sclgroup.cc/elections/about
[8] http://web.archive.org/web/20170219075544/https://sclgroup.cc/social/about
[9] http://web.archive.org/web/20170219075544/https://sclgroup.cc/social/about
[10] https://ca-political.com/services
[11] https://ca-political.com/ca-advantage

II. THE SUBJECT ACCESS REQUEST

17. On 10 January 2017, Professor Carroll requested his data from Cambridge Analytica through the company's website, by way of a section 7 DPA request ("subject access request"). His file was received on 27 March 2017. A copy of the file is exhibited ([bundle page C041 – C045]). Professor Carroll was accordingly a "data subject" relating to the data held by the Defendants for the purposes of section 1 of the Data Protection Act 1998.

18. The disclosure provided by the SCL Group/SCL Elections/Cambridge Analytica in response to the subject access request is inadequate, as detailed in the statements of the Claimant and Professor David Golumbia, as well as the expert reports.

III. THE PROFILE

19. On 27 March 2017, a letter was provided to the Claimant on Cambridge Analytica letter head, stating that Cambridge Analytica were providing the Claimant with "all of the data to which [he is] entitled under the DPA, in a Microsoft Excel Spreadsheet (.xls)" (emphasis added). The Profile, as at 27 March 2017, included information in the following categories:

- Core data
- Election returns
- Models

20. This is the only information that has been disclosed to the Claimant. However, Cambridge Analytica (and by extension SCL Group) state that they have "5,000 data points" on each profile. These data points have not been disclosed.

IV. PRE-ACTION CORRESPONDENCE

21. The Claimant served his pre-action letter before claim on 12 April 2017. I exhibit a copy of that letter ([exhibit bundle page C212 – C220]). That letter was sent first by email and by recorded delivery to the Defendants' joint address, 55 Oxford Street, London, WC1A 1BS. A number of other clients have also sent pre-action letters around the same on similar terms, including Professor David Golumbia.

22. In addition to the claims, the pre-action letter served notice under section 10 of the DPA to require the Defendants to cease unlawfully processing the Claimant's personal data.

23. A response was requested by 17 May 2017, giving the Defendants 35 days to provide a response (a generous amount of time pursuant to the practice direction on pre-action conduct and protocols). An email was served by the Defendants on 3 May 2017, in which they stated that they were "seeking legal advice" and would look to "respond to [the Claimant] fully shortly." ([exhibit bundle page C221 - 223])

24. No response was received in accordance with the protocol by 12 May 2017. A chasing letter was sent on 17 May 2017. A further chasing letter was sent on 25 May 2017. We have not, to date, received a substantive response to provide to the Court. I exhibit that correspondence ([exhibit bundle pages C224 - C226]).

25. In the absence of a satisfactory response, on 30 September 2017, the Claimant commenced a public funding campaign to support his claim. Following receipt of his target amount of funding, the Claimant instructed us to proceed with the case.

26. On 23 February 2018, we wrote to the Defendants to put them on notice that we intended to file an application pursuant to CPR 31.16 for disclosure ([exhibit bundle page C226A]). We requested the Defendant to provide a response to that letter and full disclosure by 2 March 2018. The Defendant failed to provide a satisfactory response to that letter and as such, the Claimant has been forced to apply for disclosure to properly understand his claim.

27. In the course of making that application, the Claimant instructed a number of experts to provide reports on the information that may have been missing. Those reports were commissioned from Dr David Stillwell and Professor Philip Howard, both of whom are the leading authorities in their respective fields. Those reports can be found at (bundle pages [D001 – D060]).

28. In the course of those reports, I have provided Dr Stillwell with a profile from another individual, Professor David Golumbia. Professor Golumbia's profile was provided for comparative purposes. Dr Stillwell has commented on Professor Golumbia's profile as well.

29. Professor Golumbia has also provided a short statement in support of the Claimant's application. I exhibit that statement (bundle page [B021]).

STATEMENT OF TRUTH

I believe that the facts stated in this witness statement are true.

Signed:

Named: Ravi Naik

Dated: 16 March 2018

First statement of D. Golumbia

On behalf of the Claimant

Exhibits C233 – C239

16 March 2018

IN THE HIGH COURT OF JUSTICE **Claim no.**

QUEEN'S BENCH DIVISION

B E T W E E N

DAVID CARROLL

Claimant / Applicant

- and -

(1) CAMBRIDGE ANALYTICA LTD
(2) CAMBRIDGE ANALYTICA (UK) LTD
(3) SCL ELECTIONS LTD
(4) SCL GROUP LTD

Defendants / Respondents

STATEMENT OF DAVID GOLUMBIA

I, DAVID GOLUMBIA, of 7516 Wentworth Ave, Richmond, VA, 23228, USA, STATE as follows

1. Save as set out herein, the matters set out in this witness statement are within my knowledge. Where the facts are not within my own knowledge, they are true to the best of my knowledge, information and belief and I have indicated the source of my information.

2. I am an Associate Professor at Virginia Commonwealth University. I have an academic interest in data and politics.

3. As part of that academic interest and my concern for how my data was being used, I submitted a subject access request to the Defendants. I received a file of data about me from the Defendants, which I believe to be incomplete.

4. I understand that the Claimant has submitted an application for pre-action disclosure and a claim for adequate disclosure pursuant to section 7 of the Data Protection Act 1998. I am also considering whether to bring a claim against the Defendants. However, in the interests of proportionality, I have decided not to join the Claimant's application. If the Court concludes that the Defendants' approach to requests under section 7 of the Data Protection Act is unlawful and / or that further disclosure is necessary to enable prospective claimants to assess the extent and nature of their claims, these conclusions would read across to my position. I trust that in the event that the Defendants are ordered to comply with section 7 / provide pre-action disclosure, the same information will be provided to me as part of my ongoing pre-action correspondence with the Defendants.

5. I provide the following statement to assist the Court in considering the information provided by the Defendants in response to a subject access request, and to explain my concerns about the information provided to me.

A. <u>The SAR Process</u>

6. I submitted a subject access request to the Defendants on 16 January 2017. I received a substantive response on 27 March 2017. The Profile, as at 27 March 2017, included information in the following categories (which I understand was the same as the Claimant):

- Core data – Background information, including my name, address, date of birth and voter ID.

- Election returns – All data relating to my election returns for both primary and general elections from 2000 – 2014, including who I cast my vote for.

- Models – The "models" category was a political profile of me. In summary, my political views were categorised in ten variables, ranked by order of perceived importance. In addition, my partisanship was categorised by both my registered and unregistered political preference. The profile also contained a perceived propensity to vote in the 2016 election. Finally and differently to the Claimant, I had an "Early/Absentee Vote Likelihood" rating. In essence, this "model" is a bespoke political profile of me.

7. I exhibit the disclosure I received, as well as the cover letter from Cambridge Analytica ([exhibit bundle page C233 – C239]).

B. The Profile

8. The profile was distressing. In particular, the profile suggested that the Defendant companies had modelled my political views and beliefs. There was no explanation of how this model was created nor what it was based on. It is difficult to properly asses what the profile means, as there was no guidance or explanation of how it works or what its purpose is.

9. Having considered the profile in detail, I have made some suppositions about what the conclusions of the profile / predictions mean. However, ,much is necessarily guesswork on my part, unless and until I get disclosure from the Defendants.

10. Firstly, my impression is that the "importance" rank is a synonym for what might colloquially be considered a "hot button". It is not necessarily a mark of how important I might rank each issue to myself rationally, but a rank of how much a prompt based on that issue is likely to prompt me to vote.

11. I also note that the numbers 1-10 each occur exactly once, so these are not merely measures of how important each issue is to me but how important they are relative to each other.

12. The fact that in my data "national security" is given a rank of "1" suggests to me that I am not a voter who might be moved by xenophobic advertising, such as advertising about

“terrorists" and threats from foreign actors, as these issues are interpreted through a specific political lens. I do not think it is a more objective ranking of how important as a person I think “national security” (as a more general issue) might be. Thus, the data seems to me heavily psychometric in nature; it is about what hot buttons can be pressed to influence me to action, not about the concerns that I might express to myself about how important I think each thing is.

13. If it is intended to be a “rational” list of what is important to me, in terms of what I think is important, it is inaccurate in many ways. I certainly don't think “Gun Rights” are less important than education, for example. However, this is difficult for me to know if this is an error without further disclosure. If, as the Defendants state, the purpose of the profile is to understand me on a personal level, it is quite possible that an emotionally “hot” message about guns (whether in favour of gun availability or of gun control) might in fact prompt me to action. That the Defendants know this intimate detail about me is deeply troubling, particularly where I cannot say for certain that this is how I would react.

14. Regardless of which frame is used, some of the points seem accurate within the profile. For example, it is definitely true that messages about the national debt are not going to move me very much, and I do not think it is an important issue. On the other hand, for either frame, some items seem quite inaccurate. Education and jobs are ranked very low for me, where I think they are very important. However, that again is variable on the basis of what is meant by “jobs.” That (like many items on this list) can also have a special meaning in US political discourse. I again cannot properly understand what this means without full disclosure.

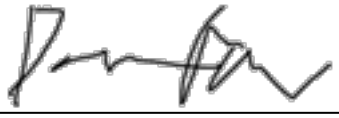
15. I also do not know what to make of the “Traditional Social and Moral Values” category, which in itself seems like a highly analytical category that must require a huge amount of data to be processed to generate the conclusion reached. To understand what this may mean I have asked myself whether my ethics line up with what some call “traditional”? I honestly don't know since I don't even know what those values are. However, I do not support what news outlets such as Fox News refer to as “traditional values” (which I would think include things like discriminating against immigrants, LGBTQ people, etc.). It is again confusing and impossible to properly decipher without further disclosure.

16. Conversely, I do think that the issue is "important" but only in the sense that I strongly oppose, and am likely to vote against candidates pressing "traditional" values of that sort.

17. Overall, it is impossible to know what the profile means unless I receive disclosure of the underlying material, an explanation of the purpose of the profile and disclosure of the recipients of the information pertaining to me.

STATEMENT OF TRUTH

I believe that the facts stated in this witness statement are true.

Signed: 

Named: David Golumbia

Dated: 16 March 2018